Repeatability and Reliability of Different Diagnostic Platforms in Normal Tear Film Evaluation

Swati Singh
L V Prasad Eye Institute
Zarin Modiwala
L V Prasad Eye Institute
Ashik Mohamed
L V Prasad Eye Institute
Sayan Basu (sayanbasu@lvpei.org)
L V Prasad Eye Institute

Research Article

Keywords: Dry eye disease, Tear film, Tear meniscus height, Lipid layer thickness, Tear break up time, NIBUT

Posted Date: September 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-864796/v1

Abstract

Purpose

To assess the repeatability and reliability of different commercially available diagnostic platforms in the objective assessment of normal tear film parameters.

Methods

Thirty healthy individuals (N = 60 eyes) had their tear film parameters (Lipid layer thickness (LLT), Tear meniscus height (TMH), Non-invasive tear break up time (NIBUT)) assessed by two observers on different occasions. The three instruments that were used are- LipiView® II, IDRA ocular surface analyzer and Oculus keratograph 5M. Bland-Altman analysis and mixed effects model with maximum likelihood estimation were used to calculate intra- and interobserver variability between the instruments.

Results

There were no significant intraobserver differences noted in LLT values measured with Lipiview, NIBUT using Oculus & IDRA, and Oculus TMH values. Between two observers, there were significant differences in LLT measurements (mean difference of 6.60; $p = 0.002$) and TMH measurements (mean difference of 0.03; $p = 0.0001$), obtained using IDRA but not for Oculus or Lipiview. Between instruments, all the measurements (LLT, NIBUT and TMH) were significantly different ($p < 0.0001$ for LLT; $p = 0.002$ for TMH; $p < 0.0001$ for NIBUT).

Conclusion

No two dry eye diagnostic platforms can be used interchangeably for the evaluation of the normal tear film. The NIBUT and LLT measurements using Oculus and Lipiview are more reliable than IDRA, and Oculus is more reliable than IDRA for TMH assessment.

Introduction

Diagnosis of dry eye disease (DED) relies on a battery of tests assessing tear film, corneal staining and tear volume.[1] As the symptomatology is similar across various etiologies of DED, the accuracy of objective testing is of paramount importance in treating the patients accordingly. Existing clinical tests such as tear break up time (TBUT) and Schirmer test have large examiner-based differences with inaccuracies in tear film height and inability to assess lipid layer. Lipid layer produced by meibomian glands maintains tear stability and its thickness correlates with gland loss.[2]

The technological development has considerably advanced for measuring tear film thickness, its constituents and stability in a non-invasive and reliable manner. The available platforms use either interferometry or videokeratographic principle. It is important to know the accuracy levels with these instruments and inter-instrumental variability of various tear film related parameters. The Diagnostic Methodology Subcommittee of the International Dry Eye Workshop has stressed the importance of developing objective methods of tear film examination which will improve comparability of measurements.[1] Currently available techniques are automated and examiner independent for measuring non-invasive tear break up time (NIBUT), lipid layer thickness (LLT) and tear meniscus height (TMH). There are different instruments available in the market and these have not been compared so far.

The current study evaluates the agreement in tear film parameters obtained using commercially available LipiView® interferometer, IDRA ocular surface analyzer and Oculus keratograph 5M.

Methods

Study Design and Participants

This study followed the tenets of the Declaration of Helsinki and was approved by the L V Prasad Eye Institute Ethics Committee (LEC_20_321). Informed consent was obtained from all subjects and/or their legal guardian(s). Thirty healthy individuals had their tear film parameters (LLT, TMH, and NIBUT) assessed by two observers on different occasions. The excluded individuals were those aged less than 18 years, contact lens wearers, those who underwent refractive surgery, those with prior ocular surgery within the past 3 months, ocular injury, or other ocular diseases such as ocular infection, allergy, or any systemic autoimmune disease.

Assessment Protocol

The ocular surface disease index (OSDI) questionnaire was filled for every participant. Two investigators (SS and ZM) operated all the instruments throughout the study, and a single measurement when labeled suitable by the respective instruments was entered into the spreadsheet. There was a 5-min interval between every measurement. Measurements were performed in the same room with relatively similar environmental conditions on the same day. The measurements were obtained by the investigators on different days between 10:00 a.m. and 16:00 p.m. in a dimly lit room with controlled temperature (21–23°C) and humidity (30–40%). One examiner (ZM) repeated the measurements in all 30 subjects for assessing the intraobserver repeatability using different instruments. The differences between two examiners were calculated for every parameter. No two readings were separate by more than 2 weeks. There was randomization of the eye to be tested first as well as the sequence of the device use. The other eye was tested 20 seconds (s) after the first while evaluating NIBUT.

Devices and Parameters Assessed

The LipiView® II Ocular Surface Interferometer (TearScience® Inc, Morrisville, NC, USA) measures the absolute thickness of the LLT by studying billion data points of the interferometric image of the tear film.

The technique of Lipiview interferometer use was similar to Eom et al.[2] The participants maintained fixation in the internal target while allowed to blink normally and video was captured for 20 seconds. The camera focus was adjusted for a clear interferometric image focused on tear film plane, and the pupil lied in the center of the live video screen. The LLT is processed in interferometric color units (ICU), where 1 ICU corresponds to approximately 1 nanometer (nm).

The IDRA Ocular Surface Analyzer (SBM Sistemi, Coburn Technologies, Inc, Italy) is a fully automated, comprehensive dry eye diagnostic system that performs non-invasive interferometry, meibography, tear meniscus height measurement and auto-NIBUT testing. The average LLT, auto-NIBUT and TMH were measured using IDRA. For LLT, interferometric principle is used i.e the surface reflection pattern of colored lipids is compared with an inbuilt grading scale. The NIBUT is measured using grids that are projected onto the cornea, and the time from the last complete blink to the first discontinuity in grid is measured as first NIBUT. It gives average NIBUT over the tested duration which was 14 seconds in our study. No manual NIBUT was calculated to avoid subjective error. TMH was measured at three points- central, medially and laterally at paracentral location. The average of three readings was taken as the final TMH.

The Oculus® keratograph 5M (K5; Oculus®, Arlington, WA, USA) is a corneal topographer, uses infrared light source to capture the placido disc image and calculate the image irregularity as NIBUT. Firstly, the image centration was done and then the participants were told to blink twice. After last blink, they were instructed to look at a fixation light and keep eyes open as long as they can. If an individual blinked before 12 seconds, the test was repeated after a relaxation time of 10 minutes.

Statistical Analysis

The statistical analysis was performed using the software STATA v14.2 (StataCorp, College Station, TX, USA). A mixed effects model applying maximum likelihood estimation was used to estimate marginal means ± standard error and make comparisons between measurements. Random intercept at the subject level was used to account for the correlation between fellow eyes of the same subject. The coefficient of repeatability was calculated as two times the standard deviation of the differences.[3] Bland-Altman analysis was performed to estimate the mean agreement and the 95% limits of agreement. A p-value of < 0.05 is normally considered statistically significant. Since there were multiple comparisons (6 in Tables 2 and <link rid="tb2">3</link>, and <link rid="tb2">3</link> in Table 4), Bonferroni correction was applied, and p-values of < 0.008 (for Tables 2 and 3) and < 0.017 (for Table 4) were considered statistically significant.

Table 2
Intra-observer agreement between two measurements (1 vs 2) of the same observer (1) in the same instrument, 95% limits of agreement and the significance level of the differences.

Comparison variable	Agreement, mean	95% limits of agreement	p-value
IDRA LLT	-1.72	-22.42 to 18.99	0.17
LipiView LLT	1.67	-28.19 to 31.52	0.45
IDRA TMH	0.03	-0.14 to 0.20	0.003
Oculus TMH	0.00	-0.18 to 0.18	0.76
IDRA BUT	0.19	-5.71 to 6.10	0.56
Oculus BUT	-0.25	-7.29 to 6.78	0.61
LLT = Lipid layer thickness; TMH = Tear meniscus height; NIBUT = Non-invasive tear break-up time			

Table 3
Inter-observer agreement first measurements of the same variable by two different observers (1 and 2) in the same instrument, 95% limits of agreement and the significance level of the differences.

Comparison variable	Instrument	Agreement, mean	95% limits of agreement	p-value
LLT	IDRA	-6.60	-54.51 to 41.32	0.002
LLT	LipiView	3.00	-64.85 to 70.85	0.33
TMH	IDRA	0.03	-0.10 to 0.16	0.0001
TMH	Oculus	0.02	-0.19 to 0.23	0.06
NIBUT	IDRA	0.57	-3.99 to 5.13	0.05
NIBUT	Oculus	0.74	-5.99 to 7.47	0.14
LLT = Lipid layer thickness; TMH = Tear meniscus height; NIBUT = Non-invasive tear break-up time				

Table 4
Inter-instrument agreement between first measurements of the same variable by the same observer (1) in two different instruments, 95% limits of agreement and the significance level of the differences.

Comparison variable	Instruments	Agreement, mean	95% limits of agreement	p-value
LLT	IDRA vs LipiView	-12.35	-64.31 to 39.61	< 0.0001
TMH	IDRA vs Oculus	-0.04	-0.23 to 0.16	0.002
NIBUT	IDRA vs Oculus	-2.80	-12.44 to 6.84	< 0.0001
LLT = Lipid layer thickness; TMH = Tear meniscus height; NIBUT = Non-invasive tear break-up time				

Results

A total of 30 healthy volunteers (N = 60 eyes) with a median age of 25 years (inter-quartile range, 23 to 32 years) participated in the study. There were 12 males and mean OSDI value was 4.2 (range, 0 to 6). None of them were contact lens users or were using any ocular medications. Slit-lamp examination was unremarkable. Tables 1 to 4 show the distribution of mean values of all parameters with intraobserver and interobserver agreement values as well as differences among various instruments. Between different diagnostic platforms, LLT, NIBUT and TMH were significantly different for same observer (Table 4).

Table 1
shows the marginal means of the measurements obtained by two observers

Device	Observer	Measurement	Variable	Mean ± standard error
IDRA	1	I	LLT	47.35 ± 1.47
		II		49.07 ± 1.52
	2	I		53.92 ± 2.85
LipiView	1	I	LLT	59.70 ± 3.58
		II		58.03 ± 3.60
	2	I		56.70 ± 3.28
IDRA	1	I	TMH	0.23 ± 0.01
		II		0.21 ± 0.01
	2	I		0.20 ± 0.01
Oculus	1	I	TMH	0.27 ± 0.02
		II		0.27 ± 0.02
	2	I		0.25 ± 0.01
IDRA	1	I	NIBUT	9.00 ± 0.32
		II		8.81 ± 0.37
	2	I		8.43 ± 0.31
Oculus	1	I	NIBUT	11.80 ± 0.55
		II		12.05 ± 0.45
	2	I		11.06 ± 0.57

LLT = Lipid layer thickness; TMH = Tear meniscus height; NIBUT = Non-invasive tear break-up time

Lipid layer thickness

The mean LLT measured with IDRA and Lipiview was 47.35 ± 1.47 micrometers (μm) and 59.70 ± 3.58 μm ($p < 0.0001$; observer 1; Table 1**).** Bland-Altman analysis showed that difference increases (0.47 mm per one mm) with increasing LLT values ($R = 0.34$; $p = 0.009$; Fig. 1B). The intraobserver coefficient of repeatability was 20.7 nm and 29.8 nm, and interobserver repeatability coefficient was 47.9 nm and 67.8 nm for IDRA and Lipiview, respectively. There were no significant differences between two measurements of the same observer (IDRA, $p = 0.17$; Lipiview, $p = 0.45$; Fig. 1). Between two observers, there were significant differences in LLT measurements (mean difference of 6.60; $p = 0.002$) obtained by IDRA but not with Lipiview (Table 3). Bland-Altman analysis showed the difference to decrease (0.90 mm per one mm) with increasing NIBUT values ($R=-0.49$; $p = 0.0001$; Fig. 1A).

Non-invasive tear break up time

The mean values of average NIBUT obtained using IDRA and K5 were 9.0 ± 0.32 s and 11.80 ± 0.55 s ($p < 0.0001$; observer 1; Fig. 1). Bland-Altman analysis showed difference to decrease (0.96 mm per one mm) with increasing NIBUT values ($R=-0.50$; $p < 0.0001$; Fig. 1C). Fifteen eyes had a K5 NIBUT less than 10 s compared to 29 eyes when measured using IDRA. The intraobserver coefficient of repeatability was 5.9 s and 7.0 s and interobserver repeatability coefficient was 4.5 s and 6.7 s for IDRA and K5, respectively There were no significant differences in NIBUT values between same observer or two observers measured using either K5 or IDRA (IDRA, $p = 0.56$; Oculus, $p = 0.61$).

Tear meniscus height

The mean TMH measured with IDRA and K5 was 0.23 ± 0.01 mm and 0.27 ± 0.02 mm that was statistically different ($p = 0.002$; observer 1; Fig. 1). Bland-Altman analysis showed difference to decrease (0.39 mm per one mm) with increasing TMH values ($R=-0.33$; $p = 0.01$; Fig. 1F). The intraobserver coefficient of repeatability was 0.17 mm and 0.18 mm and interobserver repeatability coefficient was 0.13 mm and 0.21 mm for IDRA and K5, respectively. There were no significant differences between two measurements of the same observer or between two observers using Oculus (Table 2). TMH measurements by IDRA showed a significant difference (0.03) between the first and the repeat measurements ($p = 0.003$; Fig. 1E). Between two observers, there were significant differences in TMH measurements (mean difference of 0.03; $p = 0.0001$) obtained by IDRA. Bland-Altman analysis showed the difference to increase (0.71 mm per one mm) with increasing NIBUT values ($R = 0.58$; $p < 0.0001$; Fig. 1D).

Discussion

This study reports intra- and interobserver variability for TMH, LLT and NIBUT measurements for individual instruments as well as between different instruments. There were no significant intraobserver differences noted in NIBUT measured using Oculus or IDRA, LLT (Lipiview), and Oculus TMH values. Between two observers, there were significant differences in LLT and TMH measurements, obtained using IDRA but not for Oculus or Lipiview. Between instruments, all the measurements (LLT, NIBUT and TMH) were significantly different for both observers.

The clinical tests for dry eye such as Schirmer and TBUT are considered unreliable. The variability in TBUT is up to 8 seconds between two visits and ICC was 0.29.[4] With the advent of automated measurements, the readings are more reliable now. However, there are different diagnostic platforms available in the market and there is no study that has compared the tear film parameters across them. The intraclass coefficient of repeatability of LLT using Lipiview has been noted to be 16 nm and interclass 13 nm.[3] In our study, intraclass coefficient of repeatability was 20 nm whereas interclass coefficient of repeatability was 67 nm. The sample size in study by Zhao et al. was 20 eyes compared to 60 eyes in the current study. This could be responsible for larger interobserver variation; however, the difference was not statistically significant. This variation should be kept in mind while performing clinical studies using Lipiview interferometer. The reported mean LLT values in DED patients is between 54 to 76 nm (SD,15–25) whereas 65 nm in healthy controls (33–100) that overlaps with DED patients.[5] We found mean LLT values of 48.21 nm (range, 40–92 nm) and 58 nm (range, 32–99 nm) using IDRA and Lipiview respectively, which is similar to published studies.

The average TMH value in healthy subjects using K5 has been reported as 0.27 ± 0.12 mm, 0.29 mm (range, 0.26–0.34 mm) with the intraclass correlation coefficients and coefficient of variation values of 0.914 and 16.4%, respectively for inter-individual variation.[6,7] For the intra-individual variation, the intraclass correlation coefficients and coefficient of variation values were 0.939 and 15.9%, respectively.[6] Their study did not use Bland-Altman analysis. The intraobserver and interobserver coefficient of repeatability for K5 in the current study was 0.18 and 0.21. The values did not show significant intra or interobserver variation. The differences in mean TMH using IDRA showed more intra and interobserver variation. Hence, K5 seems more reliable for measuring TMH values. NIBUT measured with the Keratograph (software version 2.73r19) in 100 healthy personnel ranged from 0.36 s to 29.0 s, with 63% of readings being < 5 s and 85% <10 s.[8] The keratograph used in the current study was the latest one, 5M. No healthy individual had value < 5 s and only 25% had value < 10 s. The reported average of NIBUT using K5 is 10.35 ± 4.2 s, 10.9 ± 3.9 s, which is comparable with our value of 11.9 ± 0.50 s.[7,9] K5 showed good repeatability for NIBUT and TMH, hence can be considered as a standard for measuring tear film quality. LLT cannot be measured as numeric value with K5, hence we did not evaluate the lipid layer status using K5.

IDRA ocular surface analyzer was introduced as a comprehensive diagnostic system for tear film analysis that can measure automated LLT, TMH, NIBUT, blink rate, meibography, pupillometry and conjunctival hyperemia grading. It had advantages over Lipiview as NIBUT and TMH could also be measured. The reported normal values of NIBUT, LLT and TMH using IDRA are 10.4 ± 2.4 s, 73.4 ± 21.9 nm, and 0.289 ± 0.16 mm, respectively.[10] However, repeatability values are not available for IDRA. We observed normal values to be 8.9 ± 0.34 s, 48.21 ± 1.49 nm and 0.21 ± 0.01 mm for NIBUT, LLT and TMH, respectively. The repeatability for TMH and LLT values was low, hence the same observer should be involved in performing dry eye measurements using the IDRA platform.

The limitation of this study was the number of eyes examined but 60 eyes were totally imaged thrice totaling 180 scans per patient and a total of 360 scans. The strength of this study is the evaluation of measurements across three dry eye diagnostic platforms, the results of which can be applied universally. The study findings suggest that no two dry eye diagnostic platforms can be used interchangeably for performing tear film studies. The NIBUT and LLT measurements using Oculus and Lipiview are more reliable than IDRA ocular surface analyzer, although a larger sample size would be required to validate the strength of the comparison.

Declarations

a) Funding: Hyderabad Eye Research Foundation (HERF)

b) Financial disclosure: None

c) Conflicts of interest: None of the authors have any conflicts of interest

References

1. Wolffsohn JS, Arita R, Chalmers R, Djalilian A, Dogru M, Dumbleton K, Gupta PK, Karpecki P, Lazreg S, Pult H. TFOS DEWS II diagnostic methodology report. Ocul Surf. 2017;15(3);539–574.
2. Eom Y, Lee JS, Kang SY, Kim HM, Song JS. Correlation between quantitative measurements of tear film lipid layer thickness and meibomian gland loss in patients with obstructive meibomian gland dysfunction and normal controls. Am J Ophthalmol. 2013;155(6):1104–10.e1102
3. Zhao Y, Tan CL, Tong L. Intra-observer and inter-observer repeatability of ocular surface interferometer in measuring lipid layer thickness. BMC Ophthalmol 2015;15;53.
4. Nichols KK, Mitchell GL, Zadnik K. The repeatability of clinical measurements of dry eye. Cornea. 2004;23(3):272-85.
5. Jung JW, Park SY, Kim JS, Kim EK, Seo KY, Kim TI. Analysis of factors associated with the tear film lipid layer thickness in normal eyes and patients with dry eye syndrome. Invest Ophthalmol Vis Sci. 2016;57;4076–83.
6. Wei A, Le Q, Hong J, Wang W, Wang F, Xu J. Assessment of Lower Tear Meniscus. Optom Vis Sci. 2016;93(11):1420-1425.
7. Tian L, Qu JH, Zhang XY, Sun XG. Repeatability and reproducibility of noninvasive keratograph 5M measurements in patients with dry eye disease. J Ophthalmol 2016;8013621.
8. Best N, Drury L, Wolffsohn JS. Clinical evaluation of the Oculus Keratograph. Cont Lens Anterior Eye. 2012;35(4):171-4.
9. Markoulli M, Duong TB, Lin M, Papas E. Imaging the Tear Film: A Comparison Between the Subjective Keeler Tearscope-Plus™ and the Objective Oculus® Keratograph 5M and LipiView® Interferometer. Curr Eye Res. 2018;43(2):155-162.

10. Vigo L, Pellegrini M, Bernabei F, Carones F, Scorcia V, Giannaccare G. Diagnostic Performance of a Novel Noninvasive Workup in the Setting of Dry Eye Disease. J Ophthalmol. 2020:5804123.

Figures

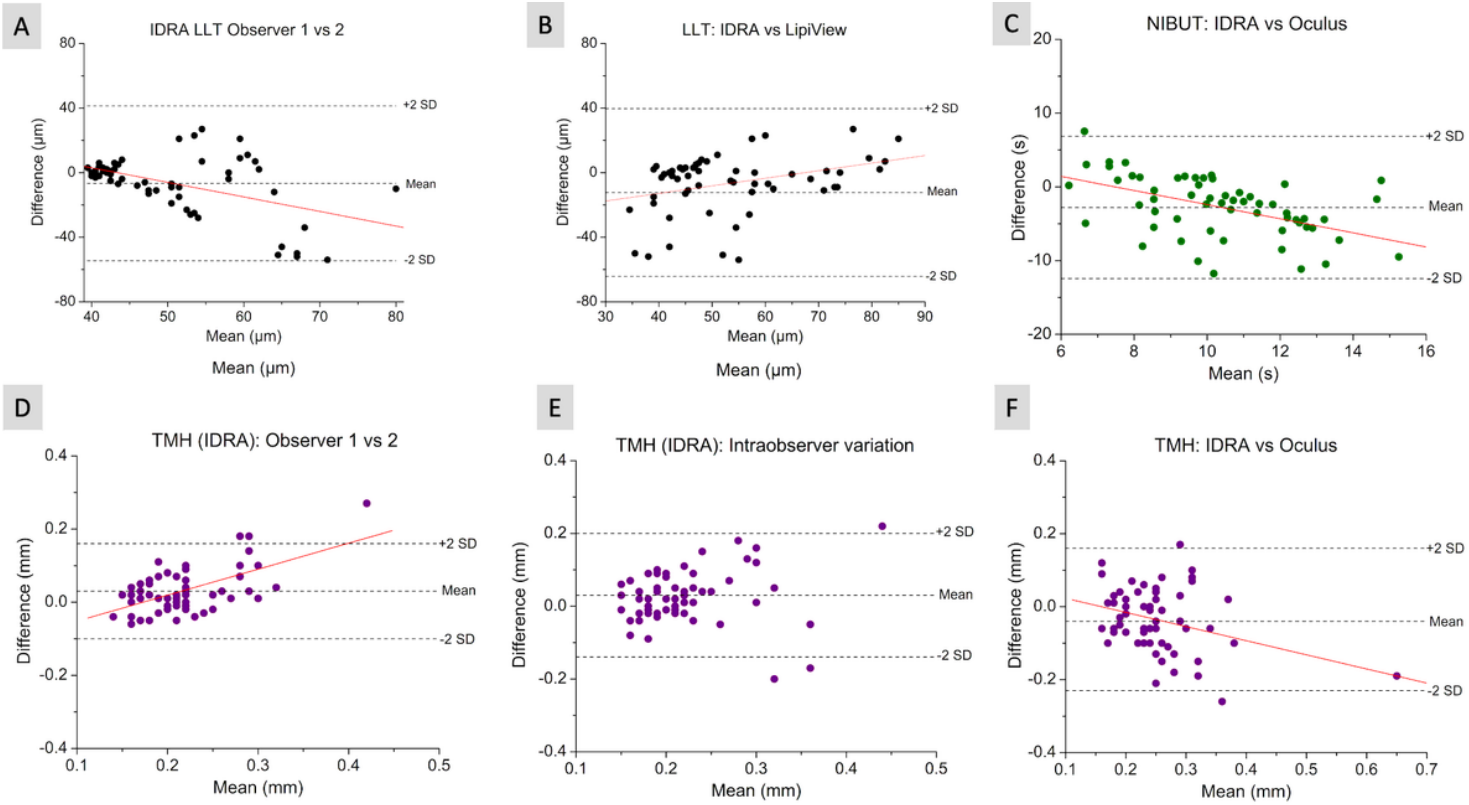

Figure 1

Scatterplots of Bland-Altman analyses. x-axis represents the average of the measurements and y-axis the difference. Mean and the 95% limits of agreement (+/-2 SD) are shown as dashed lines parallel to x-axis. Lines of best fit shows the linear relationship to be statistically significant with $p<0.05$ in panels A to D, and F.